Machine Learning-Based Data Quality Assessment and Anomaly Detection for Manufacturing Execution System Data

Saima Anjum[1]; Dr. Neelambike S.[2]

[1]Student; [2]Professor

[1,2]Department of Data Engineering, GM University, Davanagere, Karnataka, India

Publication Date: 2026/09/17

Abstract: The increasing adoption of Industry 4.0 technologies has made Manufacturing Execution Systems (MES) an essential component of modern manufacturing environments by enabling the collection, integration, and analysis of production and quality data from machines, sensors, operators, and enterprise systems. However, the effectiveness of MES-driven analytics and decision-making depends heavily on the quality, consistency, and reliability of the underlying data. Incomplete records, duplicate entries, inconsistent formats, abnormal production values, and synchronization issues can negatively affect production monitoring, process optimization, and quality management. To address these challenges, the present study proposes a Manufacturing Execution System Data Quality Assessment Framework that integrates data preprocessing, multidimensional quality evaluation, anomaly detection, standardization, and quality monitoring. The proposed workflow begins with the ingestion and normalization of MES data, followed by structural validation, data-type standardization, missing-value identification, outlier detection, temporal synchronization, and metadata-based quality flagging. Data quality is evaluated using three major dimensions: Operational Integrity, Standardization, and Optimization Impact, which are combined to calculate an overall MES data quality score. The implementation uses Python-based data processing and a dashboard-based visualization approach to identify duplicate records, defective records, production anomalies, and machine-wise defect patterns. The evaluation considers dimensions such as accuracy, completeness, consistency, timeliness, standardization, interoperability, and production alignment. The proposed framework provides a systematic and transparent approach for identifying MES data-quality issues and supports reliable manufacturing analytics and data-driven decision-making. By integrating preprocessing, quality assessment, anomaly identification, and visualization, the study provides a practical foundation for improving the reliability and usability of manufacturing data in Industry 4.0 environments.

Keywords: *Anomaly Detection; Data Preprocessing; Data Quality; Industry 4.0; Machine Learning; Manufacturing Analytics; Manufacturing Execution System.*

How to Cite: Saima Anjum[1]; Dr. Neelambike S. (2026). Machine Learning-Based Data Quality Assessment and Anomaly Detection for Manufacturing Execution System Data. *International Journal of Innovative Science and Research Technology*, 11(9), 288-294. https://doi.org/10.38124/ijisrt/26sep084

I. INTRODUCTION

The increasing adoption of Industry 4.0 technologies has significantly transformed modern manufacturing environments by enabling interconnected machines, sensors, production systems, and enterprise applications to continuously generate and exchange operational data. Manufacturing Execution Systems (MES) play an important role in this environment by integrating production, machine, operator, and quality-related information to support real-time monitoring, production control, quality management, and process optimization. The availability of reliable MES data is therefore essential for organizations seeking to improve production efficiency, identify process problems, and support data-driven decision-making. However, the growing volume and heterogeneity of manufacturing data have also introduced significant challenges related to data quality, consistency, and reliability. [1], [3], [6].

MES environments typically collect data from multiple sources, including shop-floor machines, sensors, programmable logic controllers, operator inputs, quality inspection systems, and enterprise systems such as Enterprise Resource Planning platforms. These sources may generate data using different formats, naming conventions, measurement units, and collection frequencies. As a result, manufacturing datasets can contain missing values, duplicate production records, inconsistent timestamps, incorrect data types, abnormal machine readings, and discrepancies between interconnected systems. Such data-

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26sep084

quality problems can negatively influence production monitoring, performance analysis, defect identification, predictive maintenance, and process optimization. Therefore, systematic assessment and monitoring of MES data quality are essential for ensuring that manufacturing analytics are based on reliable and trustworthy information.

Traditional data-quality approaches generally rely on predefined validation rules, manual inspection, statistical thresholds, and data-cleaning procedures to identify defective records. Although these approaches are useful for detecting clearly defined data problems, manufacturing environments continuously generate large volumes of heterogeneous data, making manual and isolated quality-checking processes difficult to maintain. Furthermore, abnormal production values may not always represent simple errors and can arise from sensor problems, synchronization issues, machine variations, operator errors, or unusual production conditions. Consequently, there is a need for systematic approaches that can combine conventional data-quality validation with automated anomaly identification and provide meaningful information about the reliability of manufacturing data.

In addition to anomaly detection, standardization represents another important challenge in MES environments. Differences in machine identifiers, production-batch formats, measurement units, timestamp representations, and process definitions can create inconsistencies between MES, enterprise resource planning, sensor, and quality-management datasets. Such inconsistencies increase preprocessing requirements and can affect cross-system analysis and production reporting. The proposed framework therefore considers Operational Integrity, Standardization, and Optimization Impact as major dimensions for evaluating MES data quality.

Based on these challenges, this study proposes a Machine Learning-Based Data Quality Assessment and Anomaly Detection Framework for Manufacturing Execution System Data. The proposed approach begins with data ingestion and normalization, followed by structural validation, data-type standardization, missing-value identification, outlier detection, temporal synchronization, and metadata-based quality flagging. The framework evaluates MES data using multiple quality dimensions and calculates an overall data-quality score. Anomaly detection is incorporated to identify abnormal production records and suspicious operational patterns, while a dashboard provides visual representations of data-quality issues, duplicate records, defective records, and machine-wise defect patterns.

➢ *The Main Contributions of this Research are Presented Below:*

- An integrated MES data-quality framework combining preprocessing, multidimensional quality assessment, anomaly detection, and visualization.
- A systematic preprocessing pipeline for identifying structural, missing-value, data-type, temporal, and outlier-related issues in manufacturing datasets.
- A multidimensional data-quality assessment approach based on Operational Integrity, Standardization, and Optimization Impact.
- An anomaly-detection mechanism for identifying abnormal production records and potential data-quality defects.
- A dashboard-based monitoring approach for presenting quality indicators, defective records, duplicate records, and machine-wise production patterns.
- A comprehensive evaluation framework considering accuracy, completeness, consistency, timeliness, standardization, interoperability, and production alignment.

The remainder of this paper is organized as follows. Section II presents the literature review related to MES data quality, manufacturing data management, anomaly detection, and Industry 4.0 data environments. Section III describes the proposed methodology, including data preprocessing, quality assessment, anomaly detection, and the overall system architecture. Section IV presents the implementation and experimental setup. Section V discusses the obtained results and evaluates the effectiveness of the proposed framework. Finally, Section VI concludes the study and presents limitations and future research directions.

II. RELATED WORK

Over the last decade, manufacturing environments have evolved significantly with the adoption of Industry 4.0 technologies, Industrial Internet of Things, smart machines, and data-driven manufacturing systems. Manufacturing Execution Systems have become an important component of these environments by integrating production, machine, sensor, operator, and quality-related information. As manufacturing organizations increasingly depend on data analytics for production monitoring and process optimization, research has progressively focused on improving the quality, consistency, reliability, and usability of manufacturing data. [1], [2], [3], [6].

➢ *Manufacturing Execution Systems and Data Quality*

Manufacturing Execution Systems integrate data generated from shop-floor equipment, sensors, operators, quality inspection systems, and enterprise applications to provide visibility into production activities. MES data supports production planning, process monitoring, quality control, inventory tracking, and performance analysis. However, integration of heterogeneous operational systems can introduce missing sensor readings, duplicate production records, inconsistent timestamps, incorrect machine identifiers, and differences in measurement units and data formats. Such issues can reduce the reliability of production analytics and affect manufacturing decisions. [1], [3].

➢ *Data Quality Assessment in Manufacturing*

Data-quality assessment approaches generally evaluate manufacturing datasets using dimensions such as accuracy, completeness, consistency, timeliness, and validity. Accuracy determines whether recorded production information correctly represents manufacturing activities, while

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26sep084

completeness measures the presence of required operational attributes. Consistency evaluates whether corresponding records remain coherent across MES, enterprise resource planning, and quality-management systems. Timeliness determines whether production events are recorded within an acceptable time period for operational decision-making. Standardization is also important because different machine vendors and production systems may use different naming conventions, measurement units, timestamp formats, and reference identifiers.

➢ *Anomaly Detection in Manufacturing Data*

Manufacturing environments continuously generate large volumes of operational data, making manual identification of abnormal records difficult. Anomalies may occur because of sensor failures, machine abnormalities, operator-entry errors, incorrect production counts, duplicate transactions, or synchronization problems between systems. Traditional approaches commonly use predefined rules, statistical thresholds, and domain-specific validation conditions. Automated anomaly detection can improve the scalability of manufacturing data-quality monitoring by identifying unusual production values and abnormal machine readings.

➢ *Machine Learning-Based Data Quality and Anomaly Detection*

The increasing availability of manufacturing datasets has created opportunities to apply Machine Learning techniques for automated data-quality analysis and anomaly detection. Machine learning-based methods can learn patterns from historical operational data and identify observations that differ from expected behavior. However, their effectiveness depends strongly on the quality of input data. Missing values, inconsistent formats, duplicate records, and noisy observations can negatively affect learning. Therefore, data preprocessing and quality assessment should be performed before applying machine learning models. [3], [4], [5].

➢ *Research Gap*

Although considerable research has addressed manufacturing data quality, several challenges remain unresolved. Existing data-quality approaches often focus on individual dimensions such as completeness, accuracy, or consistency, while fewer approaches provide an integrated assessment covering operational reliability, standardization, and the impact of data quality on manufacturing analytics. Another limitation is the continued dependence on predefined validation rules and statistical thresholds for detecting data defects. These methods can identify known errors but may have limitations when dealing with complex and previously unseen patterns in high-volume manufacturing datasets. These observations indicate the need for a systematic framework that combines data preprocessing, multidimensional data-quality assessment, machine learning-based anomaly detection, and visualization within a single MES data-quality monitoring approach.

➢ *Motivation for the Proposed Framework*

The identified research gaps motivate the development of the proposed Machine Learning-Based Data Quality Assessment and Anomaly Detection Framework for Manufacturing Execution System Data. The proposed framework combines data ingestion, normalization, structural validation, data-type standardization, missing-value identification, outlier detection, temporal synchronization, anomaly detection, and quality scoring within an integrated workflow. The framework evaluates MES data through Operational Integrity, Standardization, and Optimization Impact. By integrating data-quality assessment with automated anomaly detection and visualization, the proposed framework aims to improve the reliability, consistency, and usability of MES data in Industry 4.0 manufacturing environments.

Table 1 Summary of Existing MES Data Quality Approaches

Study Category	Advantages	Limitations
Traditional MES Data Validation	Identifies predefined data-entry and structural errors	Limited ability to detect unknown or complex anomalies
Data Quality Assessment	Evaluates accuracy, completeness, consistency, and timeliness	Often focuses on individual quality dimensions
Data Standardization	Improves interoperability between MES, enterprise, sensor and quality systems	Requires predefined standards and continuous governance
Rule-Based Anomaly Detection	Simple, transparent, and suitable for known operational constraints	May fail to identify previously unknown abnormal patterns
Statistical Anomaly Detection	Identifies unusual production values and outliers	Sensitive to thresholds and underlying data distributions
Machine Learning-Based Anomaly Detection	Can identify complex patterns and potentially unknown anomalies integrates preprocessing	Performance depends on data quality and model selection requires
Proposed Framework	multidimensional assessment, anomaly detection, quality scoring, and visualization	suitable manufacturing data and model configuration

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26sep084

III. PROPOSED METHODOLOGY

➢ *System Architecture*

The proposed framework follows a structured pipeline for assessing the quality of Manufacturing Execution System (MES) data and identifying anomalous production records. The workflow begins with the ingestion of raw manufacturing data, followed by data normalization and structural validation. The validated data is then subjected to data-type standardization, missing-value identification, duplicate-record detection, outlier analysis, temporal validation, and metadata-based quality flagging. The processed dataset is subsequently evaluated across multiple data-quality dimensions, while machine-learning-based anomaly detection is used to identify records that deviate from expected manufacturing patterns. The identified quality issues and anomaly results are integrated into a dashboard to provide a consolidated view of MES data reliability and machine-wise production defects.

➢ *MES Data Collection*

The MES dataset contains production and operational information generated during manufacturing activities. The dataset includes attributes related to production orders, machine identifiers, production quantities, defect quantities, defect rates, and defective status. These attributes provide the basis for evaluating the reliability and consistency of manufacturing records. Machine-level production and defect information is further analyzed to identify variations in manufacturing performance and records requiring additional investigation. The collected data is used as the input for preprocessing, data-quality assessment, anomaly detection, and dashboard-based visualization.

➢ *Data Preprocessing*

The raw MES dataset is first subjected to preprocessing to improve its consistency and suitability for subsequent analysis. The preprocessing workflow includes data ingestion, structural validation, data-type standardization, missing-value identification, duplicate-record detection, outlier identification, and temporal validation. Duplicate records are identified to prevent repeated observations from affecting the analysis, while missing and inconsistent values are flagged for quality assessment. Numerical production and defect attributes are validated to ensure that their values remain within meaningful operational ranges. The cleaned and validated dataset is then prepared for feature extraction, quality assessment, and machine-learning-based anomaly detection. The preprocessing stage ensures that identified anomalies are not incorrectly caused by basic formatting or data-entry problems.

➢ *Data Quality Assessment*

The quality of the processed MES data is evaluated using three major dimensions: Operational Integrity, Standardization, and Optimization Impact. Operational Integrity evaluates the reliability and correctness of production records by considering factors such as completeness, consistency, duplicate records, defect information, and valid production values. Standardization evaluates the consistency of data formats, identifiers, numerical representations, and other predefined data conventions. Optimization Impact evaluates whether the available data is sufficiently reliable for manufacturing analytics, defect monitoring, machine-level analysis, and production decision-making. The assessment of these dimensions provides a structured representation of the overall reliability of MES data.

➢ *Machine Learning-Based Anomaly Detection*

Following preprocessing and feature preparation, a machine-learning-based anomaly-detection model is applied to the processed MES dataset to identify observations that differ from normal manufacturing patterns. The selected production and operational attributes are provided as input to the anomaly-detection model. Each record is assigned an anomaly status based on its deviation from the learned pattern of normal observations. The detected anomalies are retained and flagged for further investigation rather than being automatically removed from the dataset. The machine-learning results are analyzed together with the rule-based data-quality checks to distinguish potential operational abnormalities from conventional data-quality problems. This integrated approach enables the framework to identify both predefined data issues and unusual patterns that may not be detected through conventional validation rules.

➢ *Overall Data Quality Score*

The framework combines the three major data-quality dimensions into a single overall MES data-quality score. Operational Integrity contributes 35%, Standardization contributes 35%, and Optimization Impact contributes 30% to the final score. The overall score is calculated using a weighted aggregation of the individual dimension scores:

$$DQOverall = 0.35(OI) + 0.35(S) + 0.30(OPI)$$

Where OI represents Operational Integrity, S represents Standardization, and OPI represents Optimization Impact. The resulting score provides a consolidated measure of MES data reliability while preserving the contribution of each individual quality dimension.

IV. EXPERIMENTAL SETUP

The proposed MES data-quality assessment and anomaly-detection framework was implemented using a Python-based data-processing environment and a dashboard-based visualization interface. The experimental workflow consists of dataset preparation, data preprocessing, data-quality assessment, feature preparation, machine-learning-based anomaly detection, quality-score calculation, and result visualization. The processed MES records are analyzed to identify duplicate records, defective production records, abnormal production patterns, and machine-wise defect variations. The final results are presented through a monitoring dashboard containing key data-quality and production indicators.

➢ *Dataset Description*

The experimental dataset consists of 305 MES production records containing manufacturing and operational attributes used for data-quality assessment and anomaly analysis. The dataset includes Order ID, Machine ID, Production Quantity, Defect Quantity, Defect Rate, and Defective Status. These attributes are used to evaluate production-record consistency, identify defective observations, calculate machine-wise defect quantities, and identify records that deviate from expected production patterns.

The dataset contains records associated with multiple manufacturing machines, including M01, M02, M03, and M04. The machine-level records support comparative analysis of defect quantities and help identify machines requiring further investigation.

➢ *Experimental Environment*

The implementation was carried out using Python for data processing, statistical analysis, machine-learning-based anomaly detection, and visualization. Data preprocessing and manipulation were performed using Python data-processing libraries, while the anomaly-detection procedure was implemented using a machine-learning library. A dashboard interface was developed to present the processed results and data-quality indicators.

The experimental workflow can be summarized as:

MES Dataset → Data Preprocessing → Data Quality Assessment → Feature Preparation → Anomaly Detection → Quality Score Calculation → Dashboard Visualization

➢ *Data Preprocessing*

Before applying the quality assessment and anomaly-detection procedures, the MES dataset was subjected to preprocessing. The preprocessing stage includes structural validation, data-type verification, missing-value identification, duplicate-record detection, numerical-value validation, and quality flagging. Production and defect quantities were examined to identify invalid or inconsistent observations, while duplicate records were separately identified to prevent repeated observations from affecting the analysis.

The defect rate was calculated from production and defect quantities to provide an additional indicator of production quality. The resulting processed dataset was then used for quality assessment, machine-wise analysis, and anomaly detection.

➢ *Feature Preparation for Anomaly Detection*

Relevant production attributes were selected as input features for machine-learning-based anomaly detection. Production quantity, defect quantity, defect rate, and other available numerical manufacturing attributes were considered to represent the operational characteristics of each MES record. The selected features were transformed into a suitable numerical representation before being provided to the anomaly-detection model.

Feature preparation enables the model to identify unusual combinations of production characteristics rather than relying only on predefined threshold-based rules.

➢ *Machine-Learning Experimental Configuration*

A machine-learning-based anomaly-detection approach was applied to the preprocessed MES records to identify observations that differ from normal manufacturing patterns. Production Quantity, Defect Quantity, Defect Rate, and other available numerical manufacturing attributes were used as input features. The model assigns an anomaly status to each record, and anomalous records are retained and flagged for further investigation rather than automatically removed.

The machine-learning results are compared with conventional data-quality checks, including duplicate identification, invalid-value detection, and defect-rate analysis. This combined approach supports identification of both predefined data-quality problems and unusual production patterns.

➢ *Data Quality Evaluation*

The quality of the MES dataset was evaluated using the following dimensions:

- Accuracy –evaluates whether the recorded production information correctly represents the corresponding manufacturing activity.
- Completeness – measures the availability of required values in the MES records.
- Consistency – identifies contradictory or repeated information, including duplicate records and inconsistent identifiers.
- Timeliness – evaluates whether production information is recorded within the required operational time.
- Standardization – verifies consistency in data formats, identifiers, units, and representations.
- Interoperability – evaluates the consistency of information required for integration with other manufacturing and enterprise systems.
- Production Alignment – evaluates the relationship between recorded production information and expected production outcomes.

These dimensions provide complementary information about the reliability and usability of the MES dataset.

➢ *Overall Data Quality Score*

The framework calculates an overall MES data-quality score using three aggregated dimensions: Operational Integrity, Standardization, and Optimization Impact. Operational Integrity is assigned a weight of 35%, Standardization is assigned 35%, and Optimization Impact is assigned 30%.

The overall score is calculated as:

Where:

- OI = Operational Integrity score
- S = Standardization score
- OPI = Optimization Impact score

ISSN No: - 2456-2165 https://doi.org/10.38124/ijisrt/26sep084

The resulting score provides a consolidated representation of MES data reliability while maintaining the contribution of the individual quality dimensions.

- *Dashboard-Based Evaluation*

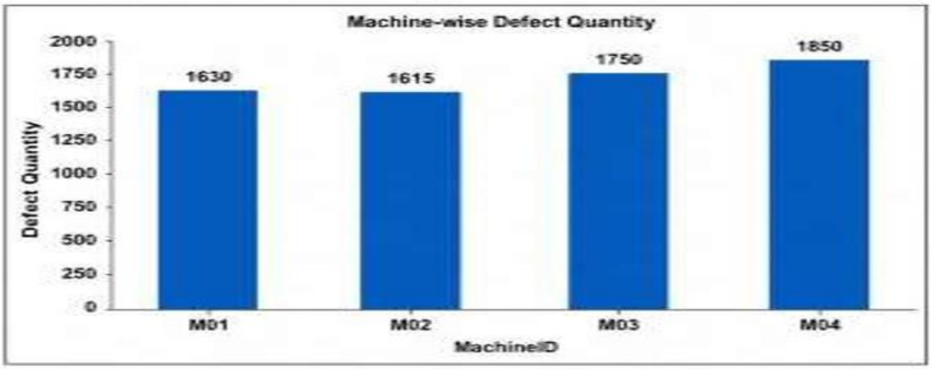
The processed results were integrated into a dashboard to provide a visual representation of MES data quality and manufacturing defects. The dashboard displays total records, duplicate records, average defect rate, total defective orders, machine-wise defect quantities, duplicate records, and defective records.

For the current dataset, the dashboard reports 305 total records, 3 duplicate records, an average defect rate of 4.48%, and 140 defective orders. Machine-wise visualization is used to compare defect quantities across M01, M02, M03, and M04. The dashboard therefore provides both an overall summary and record-level information for investigating data-quality and production-related issues.

V. RESULTS AND DISCUSSION

The implemented framework provides a systematic view of MES data quality and identifies operational records requiring further investigation. The dashboard summarizes total records, duplicate records, defective orders, average defect rate, and machine-wise defect quantities. For the current dataset, the dashboard reports 305 total records, 3 duplicate records, an average defect rate of 4.48%, and 140 defective orders.

Machine-wise Defect Graph



Duplicate Records

OrderID	MachineID	ProductionQty	DefectQty	DefectRate	Defective
1001	M03	528	37	7.01	1
1002	M04	546	34	6.23	1
1003	M01	525	27	5.14	1

Defective Records

OrderID	MachineID	ProductionQty	DefectQty	DefectRate	Defective
1001	M03	528	37	7.01	1
1002	M04	546	34	6.23	1
1003	M01	525	27	5.14	1

Fig 1 Machine-Wise Defect Graph

- *Data Quality Monitoring*

The preprocessing and validation stages enable identification of missing values, inconsistent formats, duplicate records, abnormal values, and temporal inconsistencies. These checks provide the basis for monitoring data-quality conditions across the MES dataset.

- *Anomaly Analysis*

Anomaly analysis provides a mechanism for flagging abnormal production records for investigation. Rather than automatically deleting anomalous observations, the framework preserves flagged records so that production engineers or data-quality teams can determine whether an observation represents a genuine operational event or a data-quality defect.

- *Operational Implications*

Improved MES data quality can support reliable production analytics, process optimization, quality management, and data-driven decision-making. The framework also provides a basis for identifying recurring data-quality problems and directing corrective actions toward machines, production lines, or integration points that require attention.

VI. CONCLUSION

This study presents a Machine Learning-Based Data Quality Assessment and Anomaly Detection Framework for Manufacturing Execution System data. The framework addresses common MES data-quality challenges through preprocessing, multidimensional quality assessment, anomaly identification, standardization, quality scoring, and visualization. The proposed assessment considers Operational Integrity, Standardization, and Optimization Impact, enabling manufacturing organizations to evaluate data reliability from technical and operational perspectives. The dashboard-based implementation provides a practical mechanism for monitoring duplicate records, defective records, defect rates, and machine-wise patterns. Overall, the framework provides a foundation for improving the reliability and usability of manufacturing data and supporting data-driven decision-making in Industry 4.0 environments.

LIMITATIONS

The current evaluation is based on 305 MES production records from four identified machines (M01–M04). Therefore, further validation using larger and more diverse manufacturing datasets is required to assess the generalizability of the framework. In addition, the anomaly-detection results should be evaluated using the exact model configuration and experimental outputs used in the implementation.

FUTURE WORK

- Integration of real-time streaming technologies such as Apache Kafka and Apache Flink for continuous MES data-quality monitoring.
- Development of adaptive machine learning models capable of detecting changing anomaly patterns and concept drift.
- Integration of MES with enterprise resource planning, supervisory control and data acquisition, and quality systems for automated cross-system reconciliation.
- Development of automated alerting and self-healing mechanisms for recurring data-quality defects.
- Extension of the framework to edge and cloud manufacturing environments for scalable data-quality monitoring.

REFERENCES

[1] J. Lee, B. Bagheri, and H. A. Kao, “A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems,” Manufacturing Letters, vol. 3, pp. 18-23, 2015.
[2] J. Wan, S. Tang, Z. Shu, D. Li, S. Wang, M. Imran, and A. Vasilakos, “Software-defined industrial Internet of Things in the context of Industry 4.0,” IEEE Sensors Journal, vol. 16, no. 20, pp. 7373-7380, 2016.
[3] F. Tao, Q. Qi, A. Liu, and A. Kusiak, “Data-driven smart manufacturing,” Journal of Manufacturing Systems, vol. 48, pp. 157-169, 2018.
[4] Y. Zhang, S. Ren, Y. Liu, and S. Si, “A big data analytics architecture for cleaner manufacturing and maintenance processes,” Journal of Cleaner Production, vol. 142, pp. 626-641, 2017.
[5] A. Kusiak, “Fundamentals of smart manufacturing: A multi-dimensional approach,” Annual Reviews in Control, vol. 47, pp. 214-220, 2019.
[6] J. Qin, Y. Liu, and R. Grosvenor, “A categorical framework of manufacturing for Industry 4.0 and beyond,” Procedia CIRP, vol. 52, pp. 173-178, 2016.